SEMANTIC-GUIDED FUZZY DIFFERENCE REFINEMENT FOR REMOTE SENSING SEMANTIC CHANGE DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Semantic change detection of bi-temporal remote sensing images plays an important role in urban development monitoring, environmental assessment, and land management. Recent deep learning methods have substantially improved change localization and semantic recognition. However, existing methods still struggle to preserve weak temporal evidence and make reliable decisions when real semantic changes exhibit low visual contrast. To address these challenges, we propose the Semantic-Guided Fuzzy Difference Refinement Network, termed SFDRNet, which is organized around two coordinated mechanisms: Temporal Semantic Calibration (TSC) and Uncertainty-Aware Semantic Difference Refinement (USDR). TSC constructs reliable temporal evidence by jointly modeling spatial structure and frequency statistics, and then calibrates the semantic representations of the two dates. USDR injects the calibrated semantic difference into the change representation and applies learnable fuzzy reasoning with complementary local and global evidence to refine ambiguous responses before change classification. In this way, SFDRNet first strengthens weak change evidence and then resolves the uncertainty that remains before prediction. Extensive experiments on two publicly available datasets demonstrate the effectiveness of SFDRNet. Further ablation, contrast-stratified, and uncertainty analyses validate the effectiveness of the two mechanisms in recognizing low contrast changes and refining ambiguous predictions.

1 INTRODUCTION

Semantic change detection (SCD) identifies changed regions and their land cover categories before and after change from bi-temporal remote sensing images. It supports urban monitoring, environmental assessment, and land management (Wen et al., 2021; Lv et al., 2022; Bai et al., 2023). However, a real transition may produce only slight differences in color, texture, or structure, making it difficult to distinguish from ordinary temporal variation.

Such low contrast changes are a persistent failure mode in SCD because they leave only weak temporal evidence. As shown in Fig. 1, their responses can be almost indistinguishable from ordinary variation in unchanged regions. The weak responses consequently fall near the change decision boundary and become uncertain. This overlap creates a difficult tradeoff: a conservative decision misses real changes, whereas a permissive one mistakes ordinary variations for changes. Existing methods mainly strengthen temporal comparison, multiscale context, task interaction, or semantic consistency. A few studies enhance incomplete responses through saliency or frequency representation (Tian et al., 2023; Yuan et al., 2023; Wang et al., 2025), yet subtle change evidence can still be weakened or discarded before the final decision. They also do not explicitly refine the ambiguity caused by this weak evidence. The consequence is especially severe in SCD. Once a low contrast change is removed by the binary mask, its land cover labels at both dates also disappear, so the entire semantic transition is missing from the final record. Low contrast change can therefore erase the events that the system is intended to document, making it a fundamental obstacle to reliable SCD.

To address this problem, we propose the Semantic-Guided Fuzzy Difference Refinement Network (SFDRNet), organized around Temporal Semantic Calibration (TSC) and Uncertainty-Aware Se-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

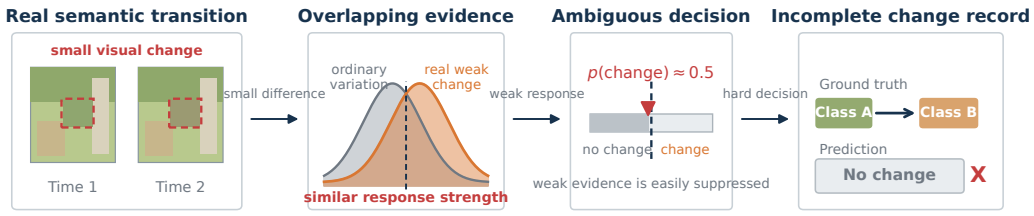

Figure 1: Low contrast change in SCD. A real transition can resemble ordinary temporal variation, leaving its response near the binary decision boundary. Suppressing this response removes both the changed region and its semantic transition. Quantitative evidence is provided in Section 4.5.

mantic Difference Refinement (USDR). As shown in Fig. 2, TSC jointly models spatial structure and frequency statistics to construct temporal evidence and calibrate the two semantic streams. USDR then conditions the change representation on their semantic difference and maps the response to learnable Gaussian memberships. Complementary local and global fuzzy responses produce a residual correction before binary classification. TSC thus strengthens weak temporal evidence, while USDR refines the ambiguity that remains. Experiments on SECOND and Landsat-SCD, together with controlled mechanism and uncertainty analyses, validate these roles.

The main contributions of this work are summarized as follows:

1. We propose SFDRNet for low contrast SCD. TSC constructs and calibrates temporal and semantic evidence, while USDR refines ambiguous change responses before prediction.
2. We introduce semantic-guided fuzzy refinement as the central technical contribution. Learnable Gaussian memberships and complementary local and global rules preserve graded evidence and generate a residual correction for binary change prediction.
3. Experiments on SECOND and Landsat-SCD provide competitive results and mechanism-specific evidence. Contrast stratification verifies the role of TSC, while parameter-matched, uncertain-pixel, and calibration analyses isolate the effect of fuzzy refinement.

2 RELATED WORK

2.1 SEMANTIC CHANGE DETECTION

Remote sensing change detection has evolved from pixel- and object-based comparison (Hussain et al., 2013) to fully convolutional Siamese networks, dense multiscale fusion, spatial–temporal attention, and Transformer context modeling (Daudt et al., 2018; Fang et al., 2022; Chen & Shi, 2020; Chen et al., 2022; Zhang et al., 2022). Semantic priors and probabilistic modeling have further broadened how binary change evidence is constructed (Zheng et al., 2024; Gan et al., 2025). SCD extends this task by jointly localizing changes and predicting their land cover categories. Early deep formulations include joint land cover mapping and change detection (Daudt et al., 2019), postclassification fusion (Xia et al., 2022), temporal symmetry and semantic–change relations (Zheng et al., 2022), and asymmetric Siamese encoding (Yang et al., 2022). Later methods improve multiscale context, task interaction, temporal correlation, or semantic consistency (Yuan et al., 2022; Cui & Jiang, 2023; Ding et al., 2022; Tian et al., 2023; Long et al., 2024; Ding et al., 2024; Li et al., 2024; Tang et al., 2024; Zhang et al., 2024). Recent designs further explore semantic alignment, semantic guidance, boundary supervision, and late-stage fusion (Fang et al., 2025; Long et al., 2025; Zhou et al., 2025; Tang et al., 2025; 2026).

These methods improve discrimination, semantic consistency, and boundary quality. TCRPN directly targets incomplete change responses through single-temporal saliency (Tian et al., 2023), while frequency-domain methods strengthen global or semantic difference features (Yuan et al., 2023; Wang et al., 2025). However, weak evidence from a real transition can still resemble ordinary temporal variation near the decision boundary between the changed and unchanged classes. These approaches do not explicitly retain a graded response and refine its ambiguity before binary prediction. Once a weak response is suppressed as background, later semantic reasoning cannot recover

the missing transition. We address this gap by first strengthening subtle temporal evidence and then refining the remaining ambiguity.

2.2 Fuzzy Feature Refinement

Fuzzy sets represent gradual membership instead of forcing an immediate binary assignment (Zadeh, 1965). Adaptive neuro-fuzzy systems further show that membership functions and inference rules can be learned from data (Jang, 1993). More recently, learnable Gaussian memberships and complementary rules have been used to describe weak visual evidence from local and image-level statistics (Zhang et al., 2026). Inspired by this principle, we condition change features on calibrated bi-temporal semantic differences before fuzzy reasoning. The resulting memberships are intermediate feature responses rather than calibrated prediction probabilities, and their residual output directly refines the feature used for the binary change mask. We separately assess output uncertainty through change probabilities and calibration, and use a parameter-matched convolutional control to isolate the effect of fuzzy refinement.

3 Proposed Method

3.1 Overall Architecture

Figure 2 shows the SFDRNet pipeline. Given registered images I_1 and I_2, a shared ResNet-34 encoder (He et al., 2016) extracts low- and high-level features F_t^l and F_t^h for $t \in \{1, 2\}$. TSC produces a decoder change feature X_c and calibrated semantic features $\widehat{S}_1$ and $\widehat{S}_2$. USDR conditions X_c on their semantic difference and returns the refined feature X_f. A binary head predicts the change mask from X_f, and two semantic heads predict land cover from $\widehat{S}_1$ and $\widehat{S}_2$. At inference, a semantic transition is retained only where the binary head predicts change.

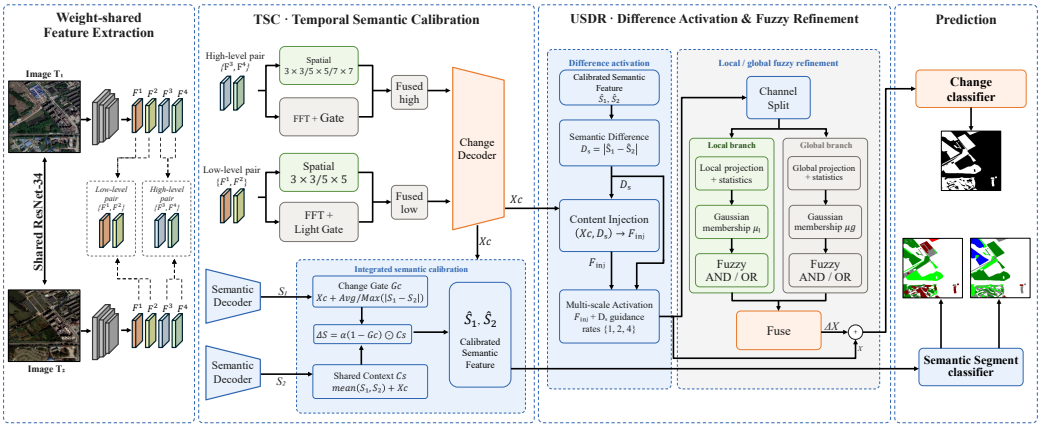

Figure 2: Pipeline of SFDRNet. TSC jointly constructs temporal evidence and calibrates the two semantic streams. USDR injects their semantic difference into the change stream and refines ambiguous responses through local and global fuzzy reasoning. The binary change mask controls the final semantic predictions at both dates.

3.2 Temporal Semantic Calibration

TSC first constructs temporal evidence from spatial structure and frequency statistics, then controls information sharing between the two semantic streams.

Temporal evidence construction. For a feature level $q \in \{l, h\}$, let $D^q = \phi_d^q(|F_1^q - F_2^q|)$ denote the projected temporal difference. We summarize its spatial structure with a multiscale mapping $\mathcal{S}_q$ and its frequency distribution with the amplitude of the 2D Fourier transform $\mathcal{F}$. Their joint guide

and the resulting temporal mixture are

$$\begin{aligned} a_f^q &= \sigma(f_q(\mathrm{GAP}(|\mathcal{F}(\mathcal{P}_q(D^q))|))), \\ G^q &= \mathcal{S}_q(D^q) + \lambda_q(D^q \odot a_f^q), \\ w^q &= \sigma(g_q([F_1^q, F_2^q, G^q])), \\ T^q &= w^q \odot F_1^q + (1 - w^q) \odot F_2^q, \end{aligned} \tag{1}$$

where f_q and g_q are learned mappings, $\mathcal{P}_q$ is level-dependent pooling, and λ_q controls the frequency contribution. The spatial mapping uses receptive fields $\{3, 5, 7\}$ at the high level and $\{3, 5\}$ at the low level. The resulting high- and low-level mixtures form residual updates of the temporal mean and $[F_1^l, F_2^l]$, respectively, and are decoded into X_c. We set $\lambda_h = 1$ and learn λ_l to introduce frequency evidence more conservatively at the low level.

Temporal semantic calibration. Two semantic decoders first produce S_1 and S_2. Directly sharing information between them can erase a true transition, so TSC estimates where shared context is appropriate. Let $D_s^0 = |S_1 - S_2|$ and $\overline{S} = (S_1 + S_2)/2$. The calibrated features are

$$\begin{aligned} c &= \sigma\big(r_\theta([X_c, \mathrm{Avg}_c(D_s^0), \mathrm{Max}_c(D_s^0)])\big), \\ Q &= q_\theta([\overline{S}, X_c]), \\ \widehat{S}_t &= S_t + \beta(1 - c) \odot Q, \qquad t \in \{1, 2\}, \end{aligned} \tag{2}$$

where Avg_c and Max_c pool over channels and β is learnable. The gate c estimates change evidence, so $1 - c$ admits shared context in stable regions and suppresses it across likely semantic transitions.

3.3 Uncertainty-Aware Semantic Difference Refinement

USDR conditions X_c on the difference between $\widehat{S}_1$ and $\widehat{S}_2$, then maps the conditioned response to learnable fuzzy sets. It thereby connects weak visual evidence to a possible semantic transition while retaining graded evidence before the binary decision.

Semantic difference conditioning. We compute $D_s = |\widehat{S}_1 - \widehat{S}_2|$ and combine its original, local, and dilated responses. A learned gate controls how much of this semantic evidence enters the change representation:

$$\begin{aligned} \widetilde{D}_s &= d_\theta([D_s, k_\theta(D_s), k_\theta^{(2)}(D_s)]), \\ X_i &= X_c + \alpha\, g_\theta(X_c, \widetilde{D}_s) \odot u_\theta(X_c, \widetilde{D}_s), \qquad 0 \le g_\theta \le 1, \end{aligned} \tag{3}$$

Here, d_θ fuses the three responses, $k_\theta^{(2)}$ has dilation two, g_θ is a sigmoid gate, and α is learnable. The fused difference generates single-channel gates at pooling rates $\{1, 2, 4\}$, which modulate X_i from coarse to fine. Their aggregation is combined with X_c and projected to obtain X_m.

Local and global Gaussian memberships. Inspired by learnable Gaussian membership modeling (Zhang et al., 2026), we refine semantic change evidence through the local and global branches in Fig. 3. We split $X_m \in \mathbb{R}^{C \times H \times W}$ by channel and project the two groups to U_l and U_g with 3×3 and 5×5 kernels. Let $\Omega_l(p)$ be the 3×3 neighborhood of position p and $\Omega_g(p)$ the complete feature map. For branch $b \in \{l, g\}$, let $m_{b,c}(p)$ and $\nu_{b,c}(p)$ denote the mean and standard deviation of projected channel c over $\Omega_b(p)$. Its normalized response and membership in fuzzy set k are

$$\begin{aligned} z_{b,c}(p) &= \frac{U_{b,c}(p) - m_{b,c}(p)}{\nu_{b,c}(p) + \epsilon}, \\ \omega_{b,c,k} &= \mathrm{softplus}(\rho_{b,c,k}) + \epsilon, \\ \mu_{b,c,k}(p) &= \exp\left[-\left(\frac{z_{b,c}(p) - a_{b,c,k}}{\omega_{b,c,k}} \right)^2 \right]. \end{aligned} \tag{4}$$

The local branch measures a response relative to its neighborhood, whereas the global branch uses the complete image. Overlapping sets retain graded support when low contrast evidence cannot be assigned from magnitude alone; the memberships are intermediate features rather than calibrated output probabilities.

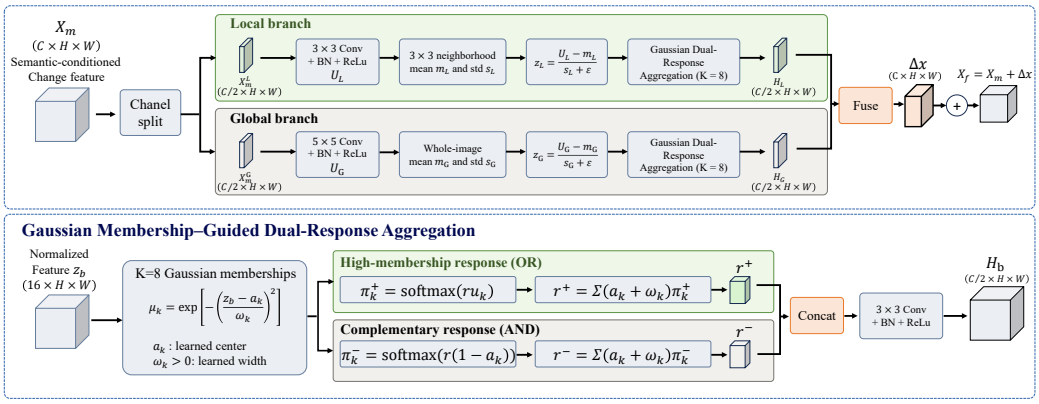

Figure 3: Fuzzy refinement in USDR. Local and global branches normalize the semantic-conditioned feature with neighborhood and image-level statistics, aggregate Gaussian memberships into complementary responses, and fuse them into a residual correction for X_m.

Rule aggregation and residual refinement. We convert the memberships into complementary rule distributions over the K fuzzy sets:

$$\begin{aligned} \pi^+_{b,c,k}(p) &= \text{softmax}_k(\tau \mu_{b,c,k}(p)), \\ \pi^-_{b,c,k}(p) &= \text{softmax}_k(\tau[1 - \mu_{b,c,k}(p)]), \\ r^{\pm}_{b,c}(p) &= \sum_{k=1}^{K} (a_{b,c,k} + \omega_{b,c,k}) \pi^{\pm}_{b,c,k}(p), \end{aligned} \tag{5}$$

where τ controls sharpness. We use $K = 8$, $\tau = 20$, $\epsilon = 10^{-6}$, and 16 projected fuzzy channels in each branch. The positive response emphasizes high-membership sets, while the complementary response retains information from the remaining sets. Writing their maps as R_b^+ and R_b^-, we obtain

$$H_b = \phi_b([U_b, R_b^-, R_b^+]), \qquad X_f = X_m + \psi([H_l, H_g]). \tag{6}$$

The 1×1 projection and normalization in ψ are initialized with zero normalization scale and bias, so USDR begins as an identity mapping and learns a residual correction without an uncertainty threshold or manually defined mask.

3.4 Prediction and Training Objective

Let h_c, h_1, and h_2 denote the change and semantic heads. Their probabilities and final semantic change maps are

$$\begin{aligned} P_c &= \sigma(h_c(X_f)), \qquad P_t = \text{softmax}(h_t(\widehat{S}_t)), \\ \widehat{Y}_t(p) &= \mathbf{1}[P_c(p) > 1/2] \arg\max_j P_t^{(j)}(p), \qquad t \in \{1, 2\}. \end{aligned} \tag{7}$$

We train the model end to end with

$$\mathcal{L} = \frac{1}{2} \sum_{t=1}^{2} \mathcal{L}_{\text{ce}}(P_t, Y_t) + \mathcal{L}_{\text{wbce}}(P_c, Y_c) + \frac{1}{|\Omega|} \sum_{p \in \Omega} \ell_{\cos}\left(P_1^+(p), P_2^+(p), 1 - 2Y_c(p)\right), \tag{8}$$

where $\mathcal{L}_{\text{ce}}$ ignores the unchanged label, $\mathcal{L}_{\text{wbce}}$ is weighted binary cross entropy, and P_t^+ is the distribution over changed semantic classes. The cosine target is $+1$ for unchanged and -1 for changed pixels. Fuzzy memberships require no auxiliary supervision.

4 Experiments

4.1 Datasets

We conduct experiments on two public SCD benchmarks. SECOND (Yang et al., 2022) contains 4,662 pairs of 512×512 high-resolution aerial images with a spatial resolution of 0.5–3 m and

covers six land-cover categories in addition to the unchanged class. We use 2,968 pairs for training and the remaining 1,694 pairs for evaluation. Landsat-SCD (Yuan et al., 2022) contains Landsat-like image pairs acquired over Tumushuke, China, from 1990 to 2020. Each image is 416×416 pixels at a spatial resolution of 30 m and is annotated with four land-cover categories—farmland, desert, building, and water—and the unchanged class. The dataset is organized into 1,431 training, 477 validation, and 477 test pairs.

4.2 Experimental Settings and Metrics

All experiments are implemented in PyTorch and performed on an NVIDIA RTX A6000 GPU. The ResNet-34 backbone is initialized with ImageNet-pretrained weights, and the remaining modules are randomly initialized. Random rotation and flipping are used for data augmentation during training, while no augmentation is applied during validation and testing. We optimize the model using SGD with a momentum of 0.9, a weight decay of 5×10^{-4}, and an initial learning rate of 0.1, which is gradually reduced by a polynomial decay schedule. The batch size is set to 8 for both datasets. The model is trained for 50 epochs on SECOND and 100 epochs on Landsat-SCD.

We report overall accuracy (OA), mean intersection over union (mIoU), semantic change F-score (F_{scd}), and separated Kappa (SeK). F_{scd} requires both correct change localization and semantic labeling, while SeK reduces the dominance of unchanged pixels when measuring semantic agreement. Higher values are better for all four metrics.

4.3 Comparison with Existing Methods

Quantitative Results. We reproduce all competing methods on the same dataset splits and evaluate them with the same metric implementation while following their original training settings. Table 1 compares our method with six representative SCD networks. On SECOND, our method achieves the highest F_{scd}, SeK, and mIoU. Relative to the strongest competing result for each metric, the improvements are 0.49, 0.52, and 0.03 percentage points, respectively. In particular, the simultaneous gains in F_{scd} and SeK indicate that the improvement is not limited to locating changed pixels, but also extends to recognizing their semantic transitions. BT-SCD obtains an OA that is 0.42 points higher than ours. However, OA is dominated by the large unchanged region, whereas F_{scd} and SeK place greater emphasis on the changed classes that define the SCD task.

On Landsat-SCD, our method ranks first on all four metrics. Compared with the strongest overall competitor, SCGNet, it improves F_{scd}, SeK, mIoU, and OA by 0.12, 0.35, 0.11, and 0.04 points, respectively. The margin is modest, but it is consistent across change recognition, semantic agreement, region overlap, and overall classification. Together with the SECOND results, this shows that the complete network remains effective on both high resolution aerial imagery and coarser Landsat imagery. These comparisons evaluate the complete model. The two mechanisms are isolated through controlled ablations on SECOND, and the behavior of fuzzy refinement is further examined on both datasets.

Table 1: Comparison on SECOND and Landsat-SCD (%). Best and second-best results are shown in bold and underlined, respectively.

	SECOND				Landsat-SCD			
Method	F_{scd}	SeK	mIoU	OA	F_{scd}	SeK	mIoU	OA
SSCD-l (Ding et al., 2022)	62.30	22.25	72.87	87.47	81.06	42.67	82.33	93.45
BiSRNet (Ding et al., 2022)	63.40	23.04	73.13	<u>88.00</u>	82.43	44.98	83.16	93.89
SCanNet (Ding et al., 2024)	63.97	23.87	73.44	87.66	86.33	53.93	86.81	95.31
DEFO-MTLSCD (Li et al., 2024)	63.57	23.79	73.70	87.81	86.14	53.83	86.89	95.31
BT-SCD (Tang et al., 2025)	64.14	<u>24.11</u>	<u>73.79</u>	**88.36**	88.23	58.40	88.41	96.04
SCGNet (Tang et al., 2026)	<u>64.20</u>	24.05	73.44	87.63	<u>90.86</u>	<u>65.02</u>	<u>90.46</u>	<u>96.81</u>
Ours	**64.69**	**24.63**	**73.82**	87.94	**90.98**	**65.37**	**90.57**	**96.85**

Qualitative Results. Representative predictions on both datasets are provided in Appendix A. They show fewer fragmented regions and more consistent semantic transitions, complementing the quantitative comparison without serving as the primary evidence.

4.4 Ablation Studies

Table 2 evaluates TSC and USDR as complete mechanisms on SECOND. The base retains the encoder, decoders, and prediction heads, and all configurations use the same split and training protocol. Because spatial and frequency operations jointly construct the temporal representation, our claim concerns TSC as a whole rather than either operation in isolation.

Table 2: Ablation of the two proposed mechanisms on SECOND (%). $\surd$ indicates that a mechanism is included. Bold denotes the best result.

Configuration	TSC	USDR	F_{scd}	SeK	mIoU	OA
Base network	–	–	63.19	23.24	73.29	87.37
TSC only	$\surd$	–	64.30	24.22	73.58	87.71
USDR only	–	$\surd$	63.37	23.51	73.58	87.84
Full model	$\surd$	$\surd$	**64.69**	**24.63**	**73.82**	**87.94**

TSC and USDR individually improve all four metrics over the base, and the full model performs best. It exceeds the base by 1.50 F_{scd}, 1.39 SeK, 0.53 mIoU, and 0.57 OA, supporting the sequence of evidence construction followed by ambiguity refinement. Table 3 then retains TSC and semantic difference conditioning while varying only the final refinement stage.

Table 3: Ablation of fuzzy refinement within USDR on SECOND (%). The convolutional control replaces fuzzy reasoning with a residual branch of similar size. Bold denotes the best result.

Refinement configuration	F_{scd}	SeK	mIoU	OA
No fuzzy reasoning	64.38	24.29	73.55	87.64
Convolutional control	64.29	24.38	73.73	87.76
Local branch only	64.33	24.22	73.48	87.60
Global branch only	64.46	24.40	73.71	87.84
Local and global	**64.69**	**24.63**	**73.82**	**87.94**

The complete fuzzy design improves over no fuzzy reasoning by 0.31 F_{scd} and 0.34 SeK. It also exceeds the convolutional control by 0.40 and 0.25 points, respectively, although their refinement branches contain 107,584 and 108,274 parameters. The global branch is beneficial alone, whereas the local branch becomes useful when paired with the image-level reference; their combination gives the best result across all metrics. These controls isolate fuzzy rule aggregation from additional capacity and support the complementary roles of the two branches.

4.5 Analysis of Low Contrast Changes

Contrast stratification protocol. We compute the smoothed mean absolute RGB difference between the two images and obtain quartile thresholds from ground truth changed pixels in the SECOND training set. The fixed thresholds 25.43, 43.55, and 69.07 partition evaluation pixels without using their performance. Table 4 compares the full model with the USDR-only configuration, thereby isolating TSC.

Table 4: Full model minus the USDR only model on SECOND pixels grouped by temporal RGB contrast. Coverage is the fraction of changed evaluation pixels in each stratum. F1 and IoU include changed and unchanged pixels in the same contrast band. Semantic accuracy is measured before binary masking, while joint SCD recall is measured after masking. All gains are percentage points.

Contrast stratum	Coverage (%)	ΔF1	ΔIoU	ΔSemantic Acc.	ΔJoint Recall
Q1, lowest	22.02	+0.76	+0.82	+3.37	+2.99
Q2	24.75	+0.57	+0.71	+2.64	+2.72
Q3	26.39	+0.41	+0.55	+2.21	+2.45
Q4, highest	26.84	+0.28	+0.39	+1.99	+2.35

Effect across contrast levels. TSC improves every stratum, with gains that decrease monotonically as contrast rises. From Q1 to Q4, the F1 gain falls from 0.76 to 0.28 points and semantic accuracy from 3.37 to 1.99 points. Thus, TSC remains useful across the range but contributes most strongly to the weakest temporal changes. The comparison evaluates the complete mechanism and does not attribute the gain separately to its spatial or frequency operations.

Link between contrast and prediction uncertainty. Predictions from the model without fuzzy reasoning show that low contrast changes are also more ambiguous before refinement. Relative to Q4, Q1 has 0.159 higher entropy, a 20.55-point higher false-negative rate, and 1.68–1.79 times as many pixels near the decision boundary. Every indicator changes monotonically across the four strata. The complete statistics and correlations are reported in Appendix B; they establish an association rather than causality and connect weak temporal evidence to the ambiguity addressed by USDR.

4.6 Analysis of Ambiguous Predictions

We use the change probability p_{wo} from the independently trained model without fuzzy reasoning to define $\Omega_m = \{i : |p_{\text{wo}}(i) - 0.5| \leq m\}$. Both models are evaluated on the same subset, preventing the full model from selecting favorable pixels. We measure output calibration with expected calibration error (ECE) (Guo et al., 2017). Table 5 combines the overall binary effect with the representative margin $m = 0.10$; complete margins and matched confident complements are reported in Appendix C.

Table 5: Full model minus no fuzzy reasoning on all pixels and on $\Omega_{0.10}$. Coverage and changes in F1 and IoU are percentages or percentage points; ECE, Brier, and NLL are unscaled and lower is better.

	All pixels			$\Omega_{0.10}$				
Dataset	ΔF1	ΔIoU	ΔECE	Coverage	ΔF1	ΔIoU	ΔBrier	ΔNLL
SECOND	+0.184	+0.234	-0.006158	5.604	+0.664	+0.564	-0.007018	-0.012255
Landsat-SCD	+0.817	+1.385	-0.003297	2.284	+7.426	+7.057	-0.028532	-0.060832

At $m = 0.10$, only 5.604% of SECOND pixels account for 99.2% of the net corrections, while 2.284% of Landsat-SCD pixels account for 66.0%. ECE also decreases on both datasets, and Brier score and NLL improve within the same fixed subsets. The gain is therefore concentrated near the original decision boundary rather than uniformly distributed over the image. Multi-margin results, confident complements, reliability diagrams, and qualitative examples are provided in Appendix C. These diagnostics concern the binary output; fuzzy memberships are not interpreted as calibrated probabilities.

Figure 4 shows two representative SECOND cases. Fuzzy refinement recovers missed regions in the binary masks, which in turn restores the corresponding semantic labels in the final maps. The aggregate behavior is quantified by Table 5; these examples only illustrate the correction pattern.

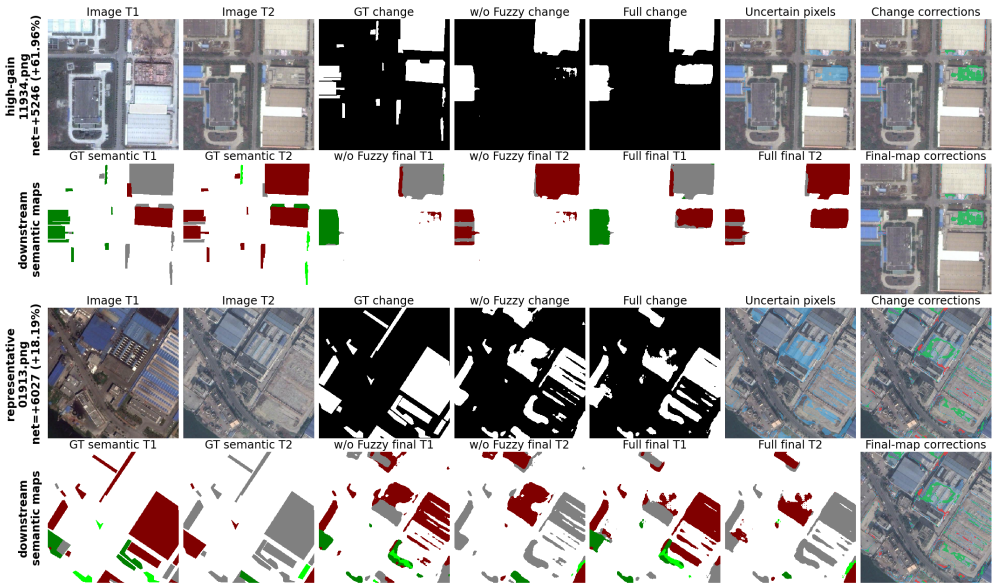

Figure 4: Representative fuzzy refinement on SECOND. The two cases show binary and semantic predictions before and after fuzzy refinement. Blue marks uncertain pixels satisfying $|p_{\mathrm{wo}} - 0.5| \leq 0.10$; green and red mark corrections and degradations by the full model.

4.7 Complexity Analysis

With paired 512×512 inputs, SFDRNet contains 26.61M parameters and requires 235.02G FLOPs. It uses 1.13M parameters and 16.37G FLOPs more than SCGNet while improving F_{scd} by 0.49 points, and remains below DEFO-MTLSCD (401.09G) and SCanNet (264.95G) in computation. Fuzzy refinement itself contains only 0.108M parameters. The full comparison is shown in Appendix D.

5 Conclusion

We presented SFDRNet for low contrast semantic change detection. TSC constructs and calibrates temporal and semantic evidence, after which bi-temporal semantic differences guide local and global fuzzy refinement before binary prediction. Experiments on SECOND and Landsat-SCD show competitive performance and consistent gains from the complete model. Contrast stratification indicates that TSC contributes most strongly to weak temporal changes, while controlled ablations and uncertain-pixel analysis show that fuzzy refinement improves responses near the decision boundary. These results support the intended sequence of evidence construction followed by ambiguity refinement.

AI use statement

Generative AI tools were used to assist with language translation and editing, manuscript organization, and the organization and presentation of existing experimental results. Generative AI was not used to generate datasets, conduct experiments, or produce numerical results; all experimental work and reported data were obtained and verified by the authors, who take full responsibility for the final content.

References

Ting Bai, Le Wang, Dameng Yin, Kaimin Sun, Yepei Chen, Wenzhuo Li, and Deren Li. Deep learning for change detection in remote sensing: A review. *Geo-spatial Information Science*, 26

(3):262–288, 2023. doi: 10.1080/10095020.2022.2085633.

Hao Chen and Zhenwei Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12(10):1662, 2020. doi: 10.3390/rs12101662.

Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. doi: 10.1109/TGRS.2021.3095166.

Fengzhi Cui and Jie Jiang. MTSCD-Net: A network based on multi-task learning for semantic change detection of bitemporal remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 118:103294, 2023. doi: 10.1016/j.jag.2023.103294.

Rodrigo Caye Daudt, Bertrand Le Saux, and Alexandre Boulch. Fully convolutional siamese networks for change detection. In *2018 25th IEEE International Conference on Image Processing*, pp. 4063–4067, 2018. doi: 10.1109/ICIP.2018.8451652.

Rodrigo Caye Daudt, Bertrand Le Saux, Alexandre Boulch, and Yann Gousseau. Multitask learning for large-scale semantic change detection. *Computer Vision and Image Understanding*, 187:102783, 2019. doi: 10.1016/j.cviu.2019.07.003.

Lei Ding, Haitao Guo, Sicong Liu, Lichao Mou, Jing Zhang, and Lorenzo Bruzzone. Bi-temporal semantic reasoning for the semantic change detection in HR remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022. doi: 10.1109/TGRS.2022.3154390.

Lei Ding, Jing Zhang, Haitao Guo, Kai Zhang, Bing Liu, and Lorenzo Bruzzone. Joint spatio-temporal modeling for semantic change detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024. doi: 10.1109/TGRS.2024.3362795.

Sheng Fang, Kaiyu Li, Jinyuan Shao, and Zhe Li. SNUNet-CD: A densely connected siamese network for change detection of VHR images. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2021.3056416.

Sheng Fang, Wen Li, Yue Song, Zhe Li, and Jianli Zhao. Rethinking semantic change detection from a semantic alignment perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–15, 2025. doi: 10.1109/TGRS.2025.3608214.

Yuhang Gan, Wenjie Xuan, Zhiming Luo, Lei Fang, Zengmao Wang, Juhua Liu, and Bo Du. Detect changes like humans: Incorporating semantic priors for improved change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–14, 2025. doi: 10.1109/TGRS.2025.3608816.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Masroor Hussain, Dongmei Chen, Angela Cheng, Hui Wei, and David Stanley. Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80:91–106, 2013. doi: 10.1016/j.isprsjprs.2013.03.006.

Jyh-Shing Roger Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3):665–685, 1993. doi: 10.1109/21.256541.

Zhe Li, Xiaoxin Wang, Sheng Fang, Jianli Zhao, Shuqi Yang, and Wen Li. A decoder-focused multi-task network for semantic change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–15, 2024. doi: 10.1109/TGRS.2024.3362728.

Jiang Long, Mengmeng Li, Xiaoqin Wang, and Alfred Stein. Semantic change detection using a hierarchical semantic graph interaction network from high-resolution remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211:318–335, 2024. doi: 10.1016/j.isprsjprs.2024.04.012.

Jiang Long, Sicong Liu, and Mengmeng Li. SMGNet: A semantic map-guided multitask neural network for remote sensing image semantic change detection. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025. doi: 10.1109/LGRS.2025.3576673.

Zhiyong Lv, Tongfei Liu, Jón Atli Benediktsson, and Nicola Falco. Land cover change detection techniques: Very-high-resolution optical images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 10(1):44–63, 2022. doi: 10.1109/MGRS.2021.3088865.

Kai Tang, Fei Xu, Xuehong Chen, Qi Dong, Yuheng Yuan, and Jin Chen. The ClearSCD model: Comprehensively leveraging semantics and change relationships for semantic change detection in high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 211:299–317, 2024. doi: 10.1016/j.isprsjprs.2024.04.013.

Xu Tang, Weihao Jiang, Jingjing Ma, Xiangrong Zhang, Fang Liu, and Licheng Jiao. SCGNet: Semantic correlation-guided network for remote sensing semantic change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–15, 2026. doi: 10.1109/TGRS.2026.3686492.

Yingjie Tang, Shou Feng, Chunhui Zhao, Yongqi Chen, Zhiyong Lv, and Weiwei Sun. A semantic change detection network based on boundary detection and task interaction for high-resolution remote sensing images. *IEEE Transactions on Neural Networks and Learning Systems*, 36(9): 17184–17198, 2025. doi: 10.1109/TNNLS.2025.3570425.

Shiqi Tian, Xicheng Tan, Ailong Ma, Zhuo Zheng, Liangpei Zhang, and Yanfei Zhong. Temporal-agnostic change region proposal for semantic change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 204:306–320, 2023. doi: 10.1016/j.isprsjprs.2023.06.017.

Renfang Wang, Kun Yang, Feng Wang, Hong Qiu, Yingying Huang, and Xiufeng Liu. Semantic change detection of bitemporal remote sensing images using frequency feature enhancement. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025. doi: 10.1109/LGRS.2025.3605910.

Dawei Wen, Xin Huang, Francesca Bovolo, Jiayi Li, Xinli Ke, Anlu Zhang, and Jón Atli Benediktsson. Change detection from very-high-spatial-resolution optical remote sensing images: Methods, applications, and future directions. *IEEE Geoscience and Remote Sensing Magazine*, 9(4): 68–101, 2021. doi: 10.1109/MGRS.2021.3063465.

Hao Xia, Yugang Tian, Lihao Zhang, and Shuangliang Li. A deep siamese postclassification fusion network for semantic change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi: 10.1109/TGRS.2022.3171067.

Kunping Yang, Gui-Song Xia, Zicheng Liu, Bo Du, Wen Yang, Marcello Pelillo, and Liangpei Zhang. Asymmetric siamese networks for semantic change detection in aerial images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022. doi: 10.1109/TGRS.2021.3113912.

Panli Yuan, Qingzhan Zhao, Xingbiao Zhao, Xuewen Wang, Xuefeng Long, and Yuchen Zheng. A transformer-based siamese network and an open optical dataset for semantic change detection of remote sensing images. *International Journal of Digital Earth*, 15(1):1506–1525, 2022. doi: 10.1080/17538947.2022.2111470.

Shiying Yuan, Ruofei Zhong, Qingyang Li, and Yaxin Dong. MFGFNet: A multi-scale remote sensing change detection network using the global filter in the frequency domain. *Remote Sensing*, 15(6):1682, 2023. doi: 10.3390/rs15061682.

Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965. doi: 10.1016/S0019-9958(65)90241-X.

Cui Zhang, Liejun Wang, Shuli Cheng, and Yongming Li. SwinSUNet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. doi: 10.1109/TGRS.2022.3160007.

Xiuwei Zhang, Yizhe Yang, Lingyan Ran, Liang Chen, Kangwei Wang, Lei Yu, Peng Wang, and Yanning Zhang. Remote sensing image semantic change detection boosted by semi-supervised contrastive learning of semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi: 10.1109/TGRS.2024.3395135.

Yang Zhang, Meibin Qi, Kunyuan Li, Xiaohong Li, and Shuo Zhuang. FLINet: Fuzzy logic for infrared small-target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 64, 2026. doi: 10.1109/TGRS.2026.3687276.

Zhuo Zheng, Yanfei Zhong, Shiqi Tian, Ailong Ma, and Liangpei Zhang. ChangeMask: Deep multi-task encoder–transformer–decoder architecture for semantic change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183:228–239, 2022. doi: 10.1016/j.isprsjprs.2021.10.015.

Zhuo Zheng, Yanfei Zhong, Ji Zhao, Ailong Ma, and Liangpei Zhang. Unifying remote sensing change detection via deep probabilistic change models: From principles, models to applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 215:239–255, 2024. doi: 10.1016/j.isprsjprs.2024.07.001.

Chenyao Zhou, Haotian Zhang, Han Guo, Zhengxia Zou, and Zhenwei Shi. A late-stage bitemporal feature fusion network for semantic change detection. *IEEE Geoscience and Remote Sensing Letters*, 22:1–5, 2025. doi: 10.1109/LGRS.2024.3507292.

A Qualitative Comparisons

Figures 5 and 6 provide full qualitative comparisons. On SECOND, SFDRNet produces more complete objects and more consistent semantic transitions. On Landsat-SCD, it reduces holes in changed regions and isolated responses in unchanged surroundings.

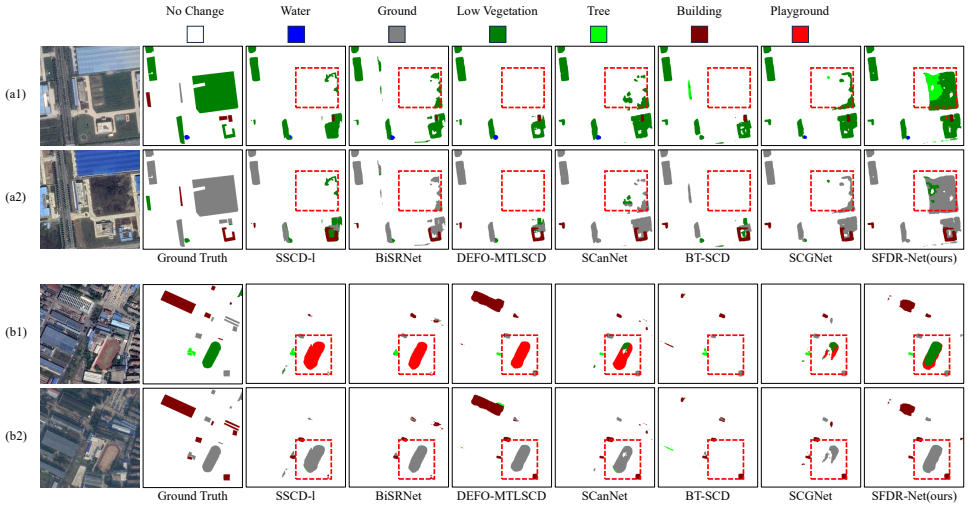

Figure 5: Qualitative comparison on SECOND. Rows (a1)–(a2) and (b1)–(b2) show two bi-temporal examples and their semantic change predictions.

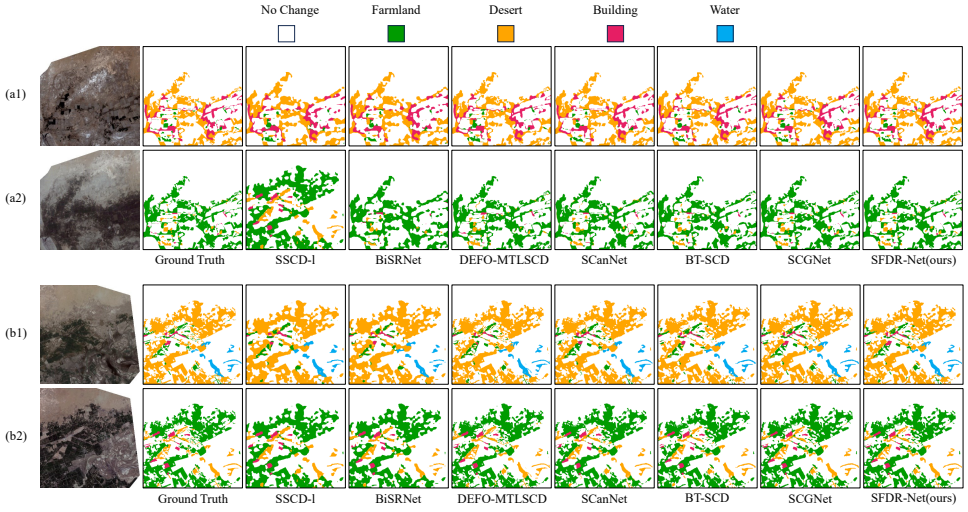

Figure 6: Qualitative comparison on Landsat-SCD. Rows (a1)–(a2) and (b1)–(b2) show two bi-temporal examples and their semantic change predictions.

B Contrast and Uncertainty Details

Table 6 reports predictions from the model without fuzzy reasoning on ground truth changed pixels of SECOND. Every indicator varies monotonically with temporal contrast. Pixel-level Pearson correlations are -0.168 between contrast and entropy, $+0.148$ between contrast and $|p - 0.5|$, and $+0.195$ between contrast and change probability.

Table 6: Relation between temporal contrast and prediction uncertainty on ground truth changed pixels of SECOND. U_m is the percentage satisfying $|p - 0.5| \leq m$.

Contrast stratum	Mean p	Entropy	$\lvert p - 0.5 \rvert$	FN rate (%)	Brier$^+$	$U_{0.05}$ (%)	$U_{0.10}$ (%)	$U_{0.20}$ (%)
Q1, lowest	0.5996	0.5994	0.3042	37.98	0.2658	7.03	14.04	28.20
Q2	0.6947	0.5285	0.3333	26.76	0.1887	5.53	11.07	22.54
Q3	0.7514	0.4720	0.3550	20.52	0.1460	4.57	9.13	18.84
Q4, highest	0.7807	0.4402	0.3671	17.43	0.1244	3.92	7.97	16.81

C Full Ambiguity and Calibration Results

Table 7: All-pixel binary results. F1, IoU, and accuracy are percentages; lower ECE and MCE are better.

Dataset	Model	F1	IoU	Accuracy	ECE	MCE
SECOND	No fuzzy reasoning	74.418	59.258	89.663	0.031092	0.151236
SECOND	Full model	**74.602**	**59.492**	**89.908**	**0.024934**	**0.123863**
SECOND	Difference	+0.184	+0.234	+0.245	-0.006158	-0.027373
Landsat-SCD	No fuzzy reasoning	90.931	83.370	96.658	0.007568	0.073450
Landsat-SCD	Full model	**91.748**	**84.755**	**96.989**	**0.004270**	**0.042747**
Landsat-SCD	Difference	+0.817	+1.385	+0.331	-0.003297	-0.030703

Table 8: Full model minus no fuzzy reasoning on uncertain subsets Ω_m. Negative Brier and NLL differences are better. ΔN counts corrected minus degraded labels.

Dataset	m	Coverage (%)	ΔF1	ΔIoU	ΔAcc.	ΔBrier	ΔNLL	ΔN
SECOND	0.05	2.779	+1.642	+1.373	+5.770	-0.007046	-0.012166	+712,227
SECOND	0.10	5.604	+0.664	+0.564	+4.323	-0.007018	-0.012255	+1,075,791
SECOND	0.20	11.692	-0.055	-0.048	+2.465	-0.006896	-0.012852	+1,279,624
Landsat-SCD	0.05	1.134	+9.261	+8.607	+11.578	-0.029158	-0.061679	+108,345
Landsat-SCD	0.10	2.284	+7.426	+7.057	+9.550	-0.028532	-0.060832	+180,060
Landsat-SCD	0.20	4.692	+4.834	+4.873	+6.515	-0.026704	-0.058994	+252,317

On SECOND, the $m = 0.20$ uncertain subset shows small F1 and IoU decreases despite improved accuracy, Brier score, and NLL. This indicates that the discrimination gain is concentrated in a narrower boundary region. Landsat-SCD remains positive at all margins. Calibration improves more broadly: ECE decreases by 19.8% on SECOND and 43.6% on Landsat-SCD, while MCE decreases by 18.1% and 41.8%, respectively.

Table 9: Full model minus no fuzzy reasoning on matched confident complements $\overline{\Omega}_m$.

Dataset	m	Coverage (%)	ΔF1	ΔIoU	ΔAcc.	ΔBrier	ΔNLL	ΔN
SECOND	0.05	97.221	-0.057	-0.074	+0.086	-0.001214	-0.003375	+372,157
SECOND	0.10	94.396	-0.162	-0.219	+0.002	-0.001042	-0.003107	+8,593
SECOND	0.20	88.308	-0.192	-0.275	-0.050	-0.000646	-0.002397	-195,240
Landsat-SCD	0.05	98.866	+0.500	+0.865	+0.202	-0.001955	-0.006477	+164,614
Landsat-SCD	0.10	97.716	+0.286	+0.504	+0.115	-0.001649	-0.005847	+92,899
Landsat-SCD	0.20	95.308	+0.063	+0.115	+0.026	-0.001060	-0.004548	+20,642

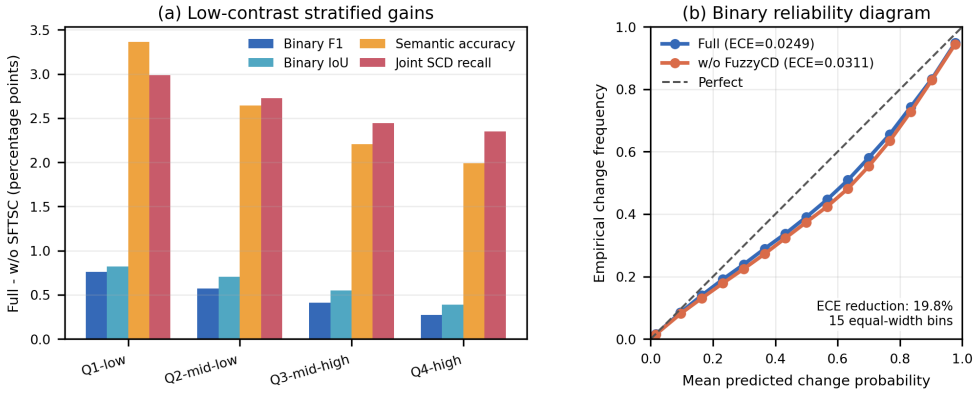

Figure 7: Mechanism-specific diagnostics on SECOND. (a) TSC gains across contrast strata. (b) Reliability diagrams for the full model and no fuzzy reasoning.

D Complexity Details

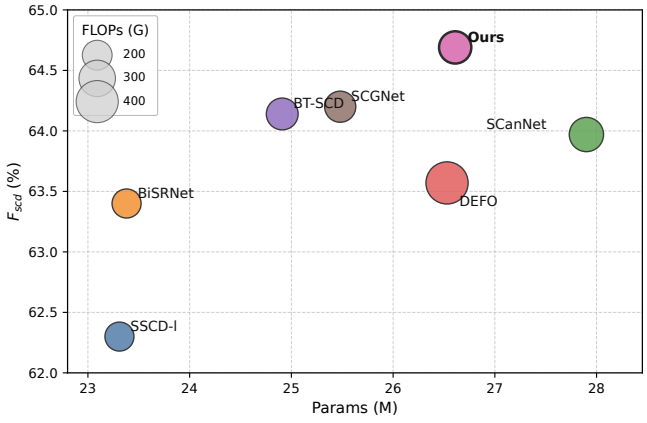

Figure 8: Accuracy-complexity comparison on SECOND. Bubble area is proportional to FLOPs.